Universidad
Popular del Cesar

Reconocimiento de Patrones

Version 2023-2

MÁQUINAS DE SOPORTE VECTORIAL

Dr. José Ramón Iglesias
DSP-ASIC BUILDER GROUP
Director Semillero TRIAC
Ingenieria Electronica
Universidad Popular del Cesar

Objetivos

Identificar el método de funcionamiento de una máquina de soporte vectorial.

Reconocer el principio de funcionamiento del truco del kernel.

Usar una SVM en el problema de clasificación binaria.

Que son:

Son un conjunto de métodos de aprendizaje supervisado para la clasificación y regresión.

Viendo los datos de entrada como conjuntos de vectores en un espacio n-dimensional, una SVM construirá un híper plano de separación en ese espacio, que maximiza el margen entre los conjuntos de datos.

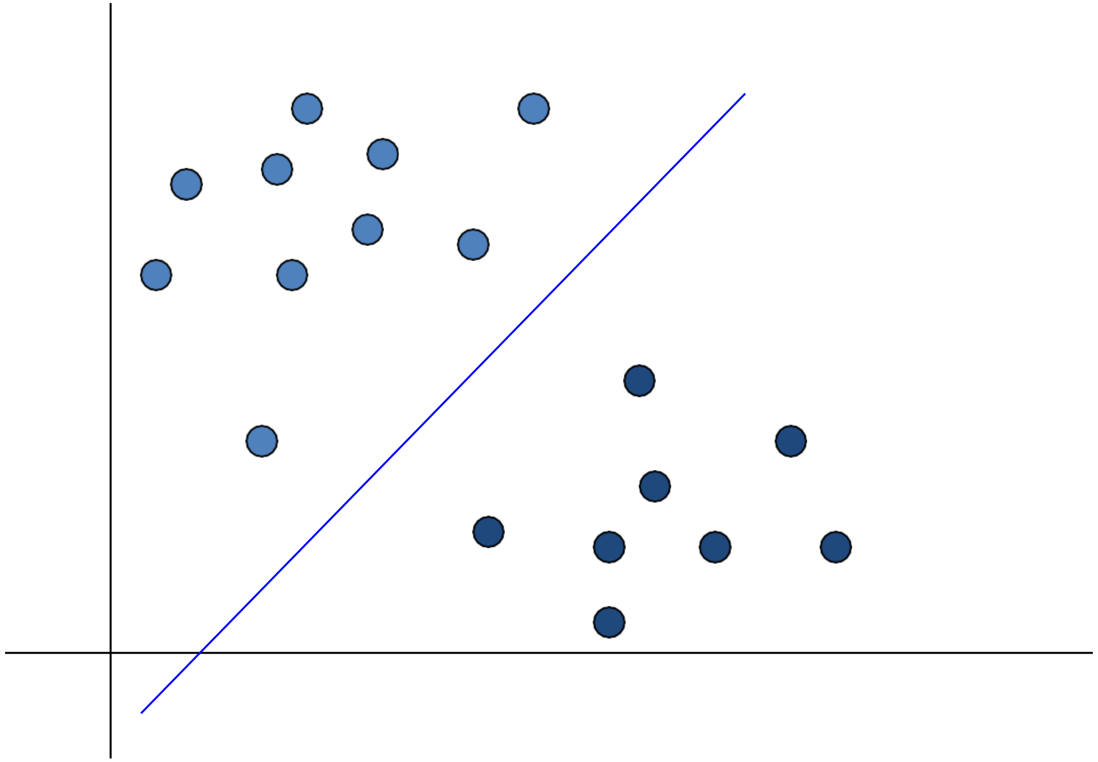
Se hicieron populares por su éxito en reconocimiento de dígitos manuscritos

Historia:

https://scholar.google.com/citations?user=U_IVY50AAAAJ&hl=en

https://scholar.google.com/citations?user=vtegaJgAAAAJ&hl=en

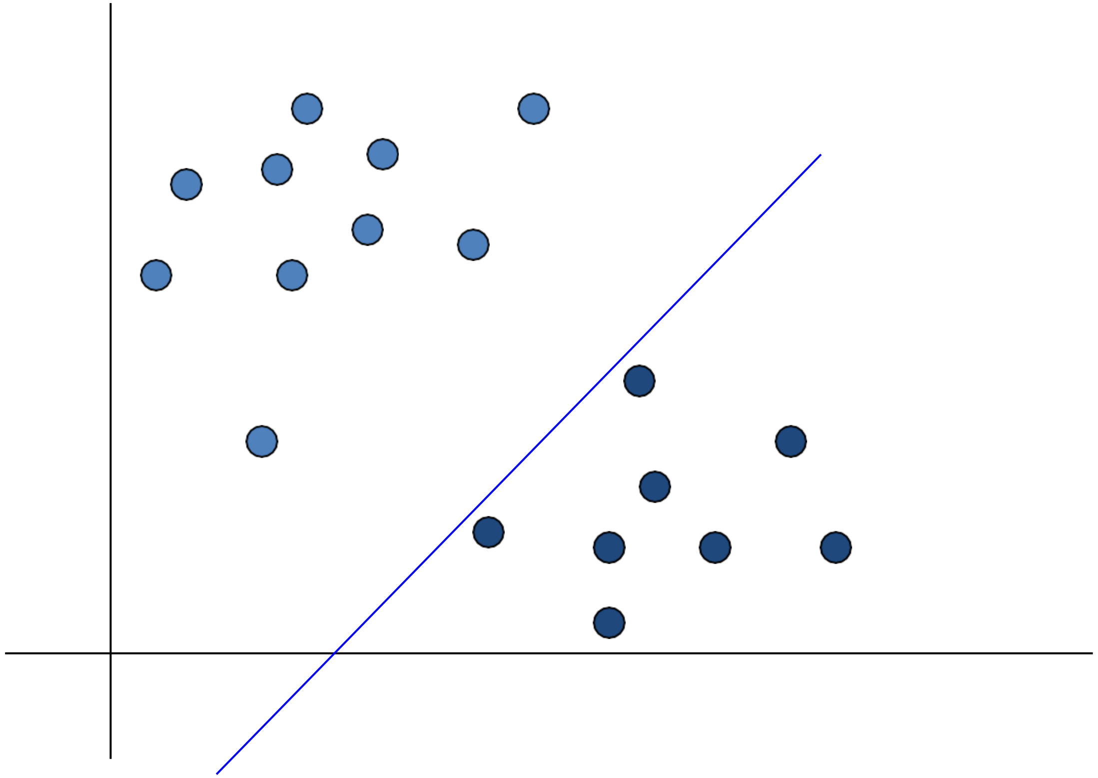
Idea:



Por ejemplo:

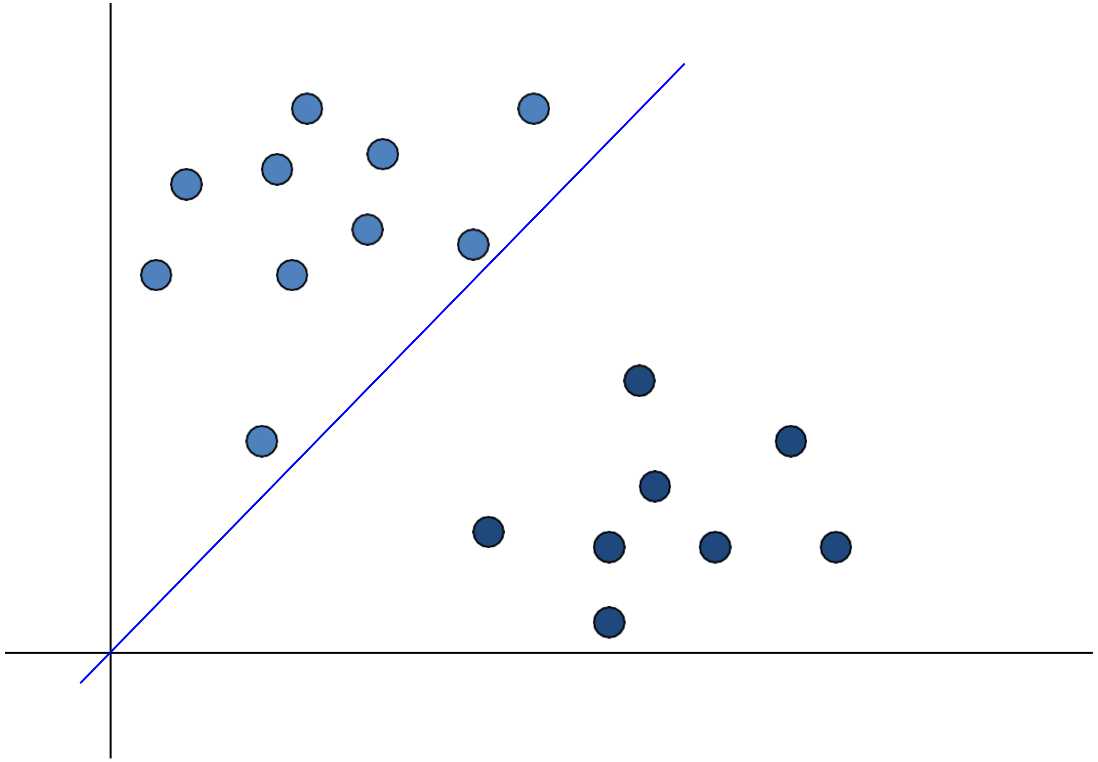
Separar en este caso dos clases con una línea recta

Idea:



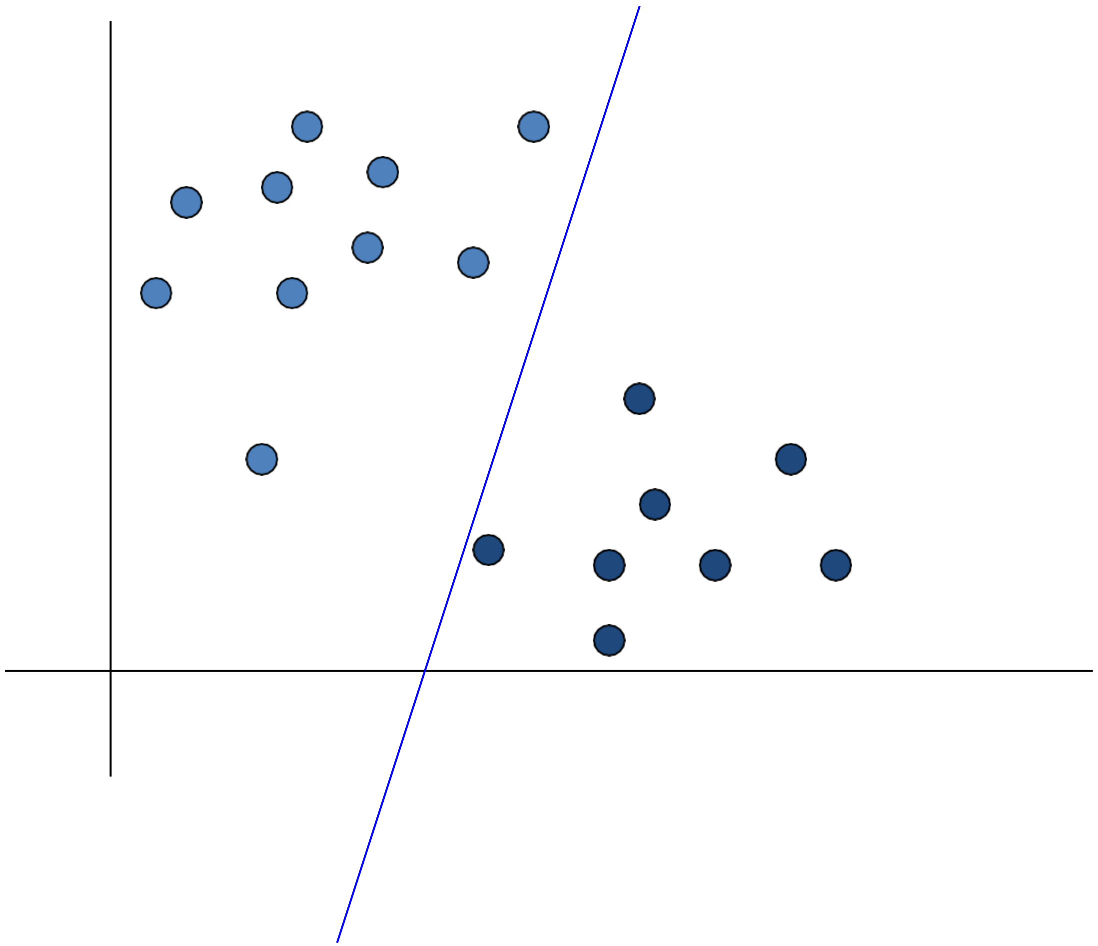
Se puede
así

Idea:



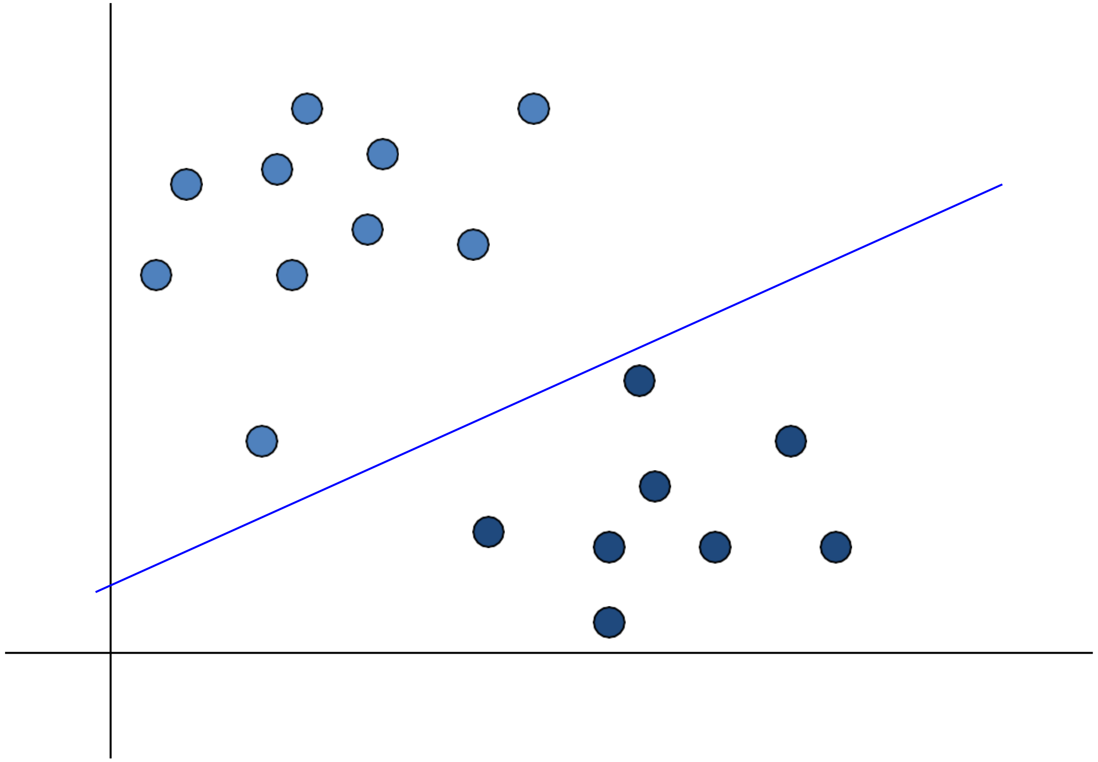
Se puede
poner
también así

Idea:



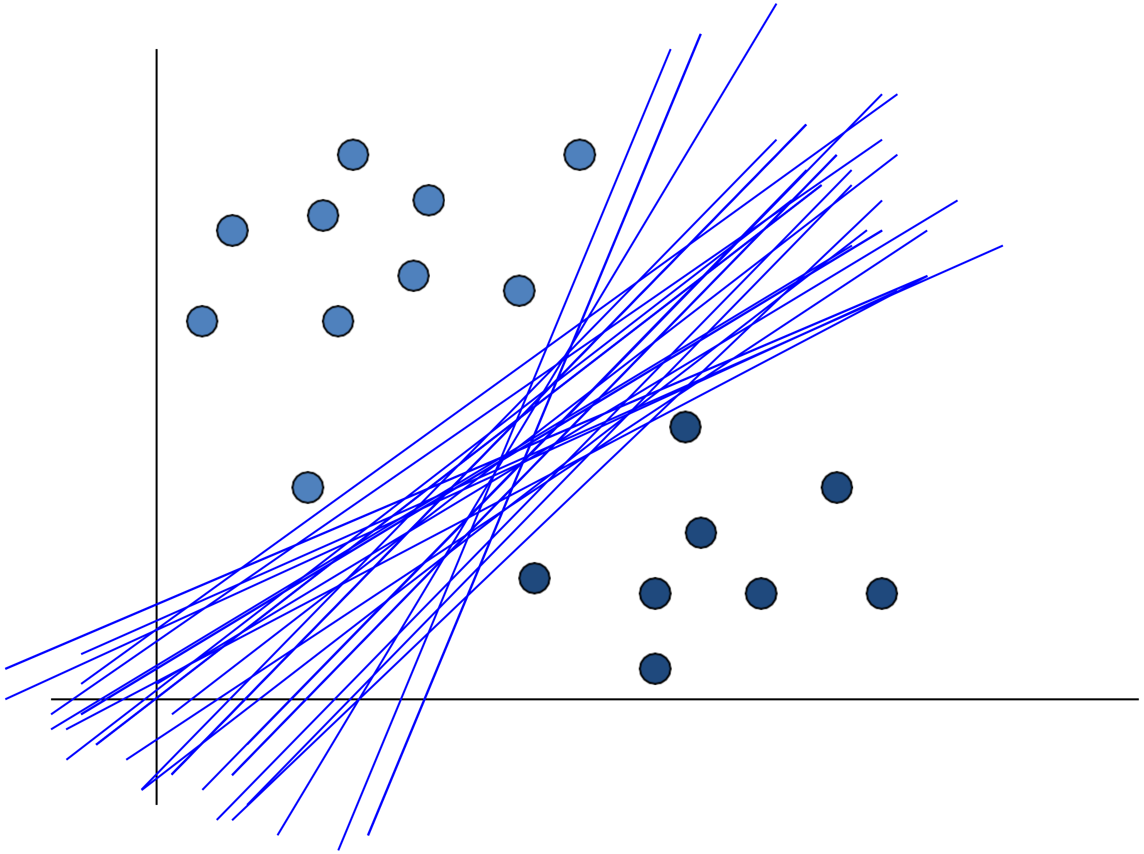
O así

Idea:



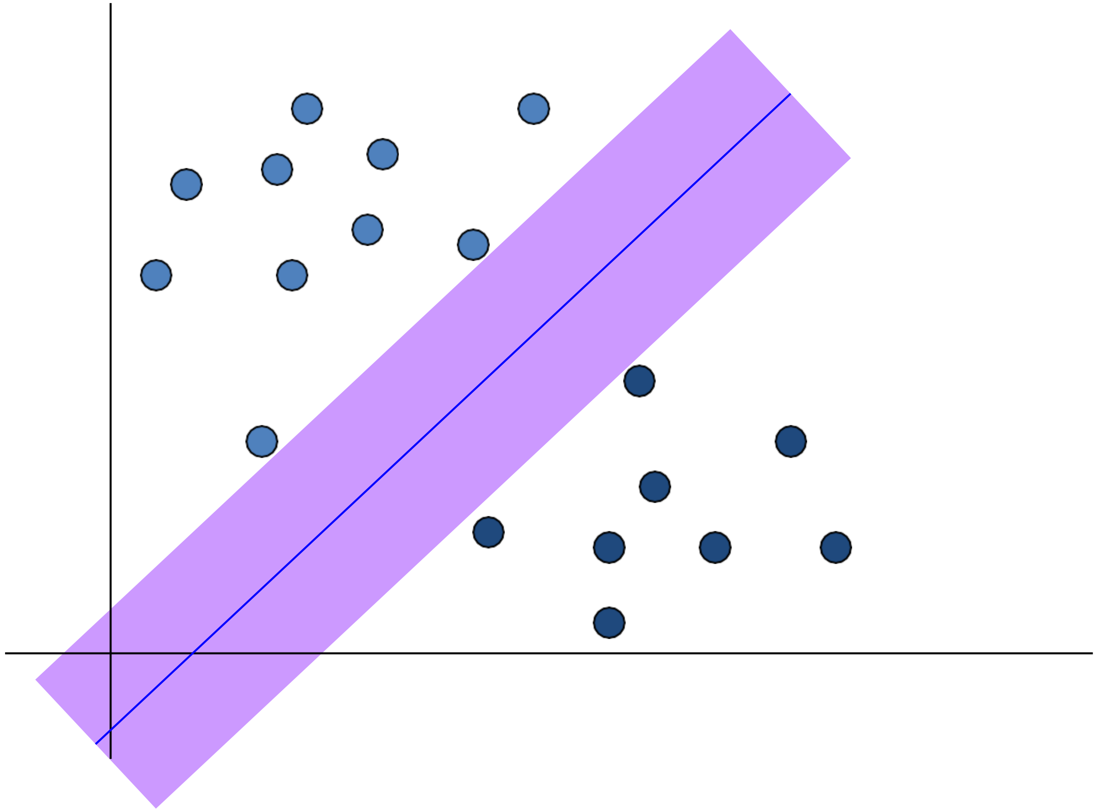
O así

Idea:



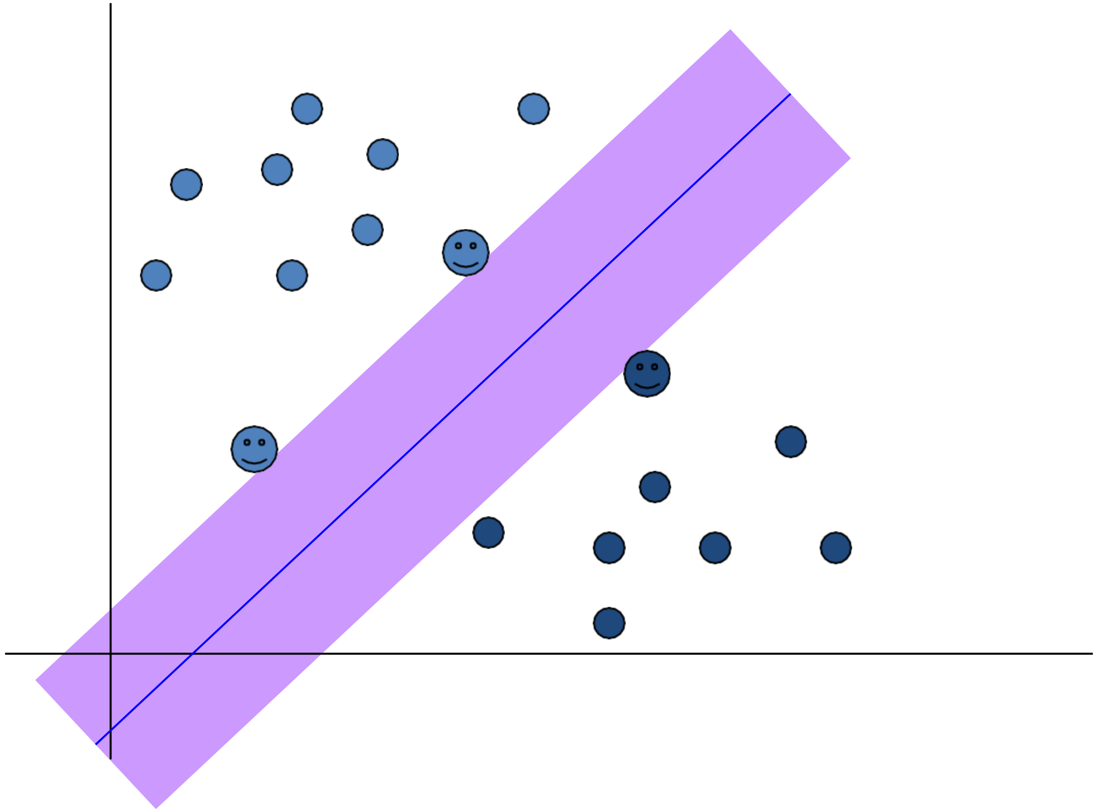
?

Idea:



La línea que
tenga un mayor
margen será la
escogida para
clasificar

Idea:



Los limites que toca este margen son llamados:

VECTORES DE SOPORTE

Maquinas de soporte vectorial lineales

Caso separable

Dadas m observaciones:

Con un vector de características $x_{\mathrm{i}} \in R^n, i = \{0 \ldots m-1\}$

Y unas etiquetas $y_i \in \{+1, -1\}$

Maquinas de soporte vectorial lineales

Supóngase que se tiene un hiperplano que separa unas muestras etiquetadas (+1) de las otras (-1). Los puntos x en el hiperplano satisfacen $\boldsymbol{w} \cdot \boldsymbol{x} - b = 0$.

w es un vector normal al híper plano.

$\|\mathbf{w}\|$ es la norma euclidiana de **w**

$\frac{|-b|}{\|\boldsymbol{w}\|}$ es la distancia perpendicular del hiperplano al origen.

Maquinas de soporte vectorial lineales

Supongamos que el conjunto de datos satisface:

$(\mathbf{w}\ "\ \mathbf{x}_i) - b \geq 1\ si\ y_{\mathrm{i}} = 1$ $(\mathbf{w}\ "\ \mathbf{x}_i) - b = 1$ con una distancia al origen de $\frac{|\&b\&(|}{||\mathbf{w}||}$

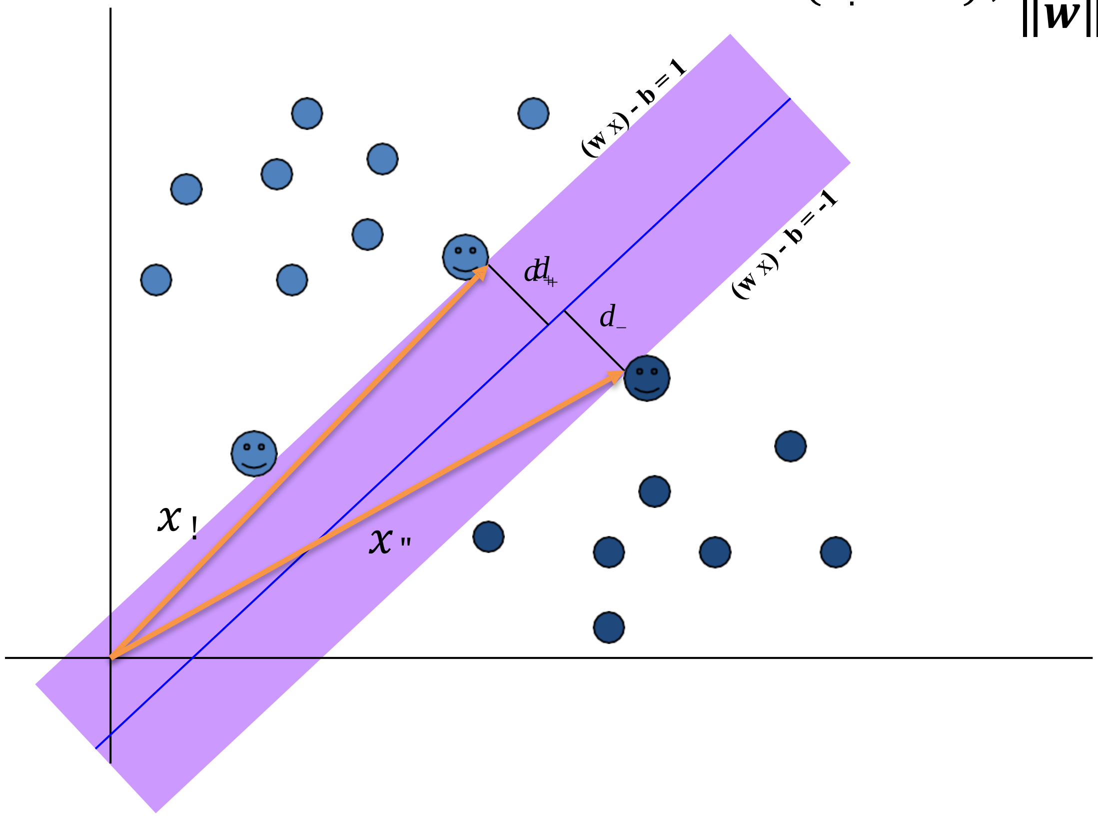
$(\mathbf{w}\ "\ \mathbf{x}_i) - b \leq -1\ si\ y_{\mathrm{i}} = -1$ $(\mathbf{w}\ "\ \mathbf{x}_i) - b = -1$ con una distancia al origen de $\frac{|(\&b|}{||\mathbf{w}||}$

Multiplicando y_{i} a ambos lados de ambas ecuaciones pero simplificando solo en la derecha:

$$y_{\mathrm{i}}((\mathbf{w}\ \$\mathbf{x}_i) - b) - 1 \geq 0$$

Maquinas de soporte vectorial lineales

$$(x_1 - x_2) \cdot \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$$



Por lo que las distancias d_+ y d_- están dadas por :

$$d_+ = d_- = \frac{|1|}{\|\mathbf{w}\|}$$

Y la suma de las dos será :

$$\frac{2}{\|\mathbf{w}\|}$$

Maquinas de soporte vectorial lineales

Se busca maximizar $\frac{2}{\|\mathbf{w}\|}$ que es lo mismo que minimizar $\|\mathbf{w}\|^2$

Sujeto a $y_i(\mathbf{w} \cdot x_i + b) \geq 1$

Este es un problema que en optimización se conoce como:

programación cuadrática: https://en.wikipedia.org/wiki/Quadratic_programming

Puede ser resuelto mediante multiplicadores de Lagrange (α_i).

$$L_P \equiv \tfrac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i + b) + \sum_{i=1}^{l} \alpha_i$$

https://es.wikipedia.org/wiki/Multiplicadores_de_Lagrange#Ejemplo_2

Maquinas de soporte vectorial lineales

También puede ser hallado en términos de los vectores de soporte, mediante la representación dual, Haciendo que los gradientes de Lp respecto a **w** y b sean cero, minimizando L_D con las restricciones (1) y (2)

$$(1)\quad \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$(2)\quad \sum_i \alpha_i y_i = 0$$

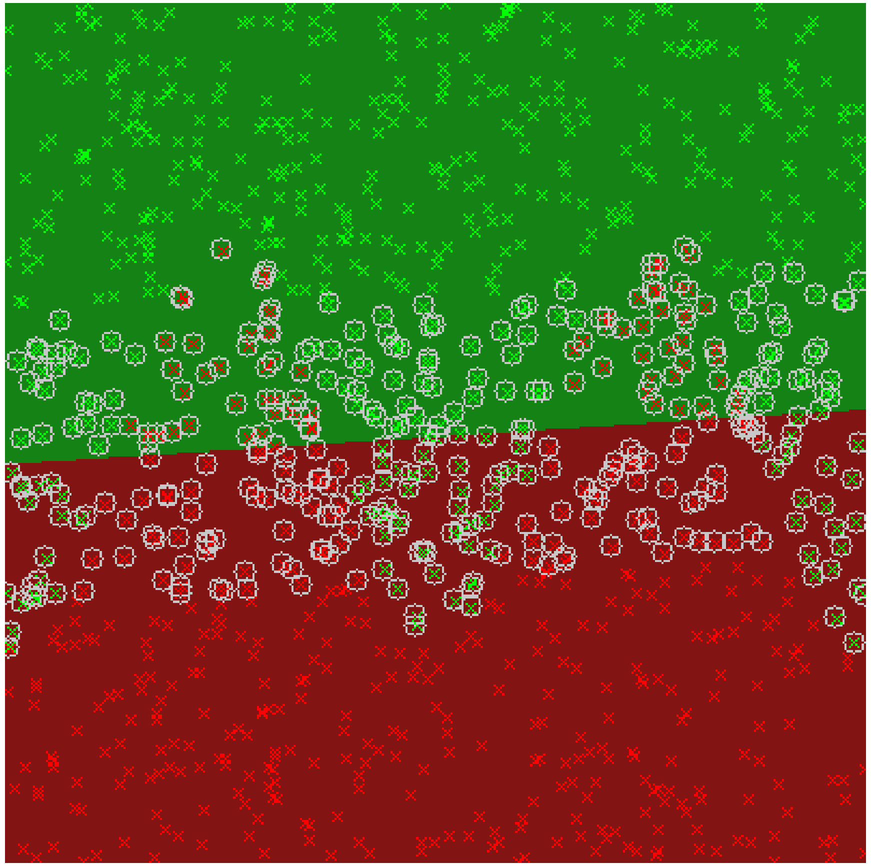
$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_i \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$$

- Los **x**i con ai no nulos son llamados vectores de soporte (*support vectors*) (VS)

Maquinas de soporte vectorial lineales

Caso no separable

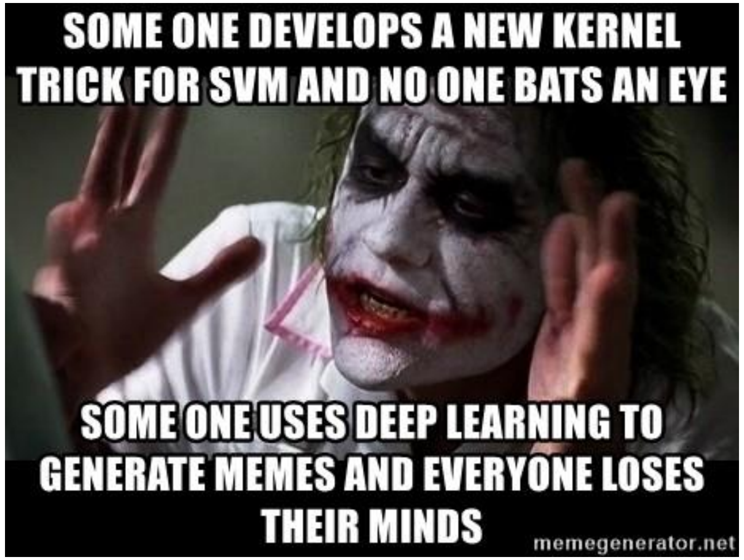
En este caso se permite cierto error en la clasificación.



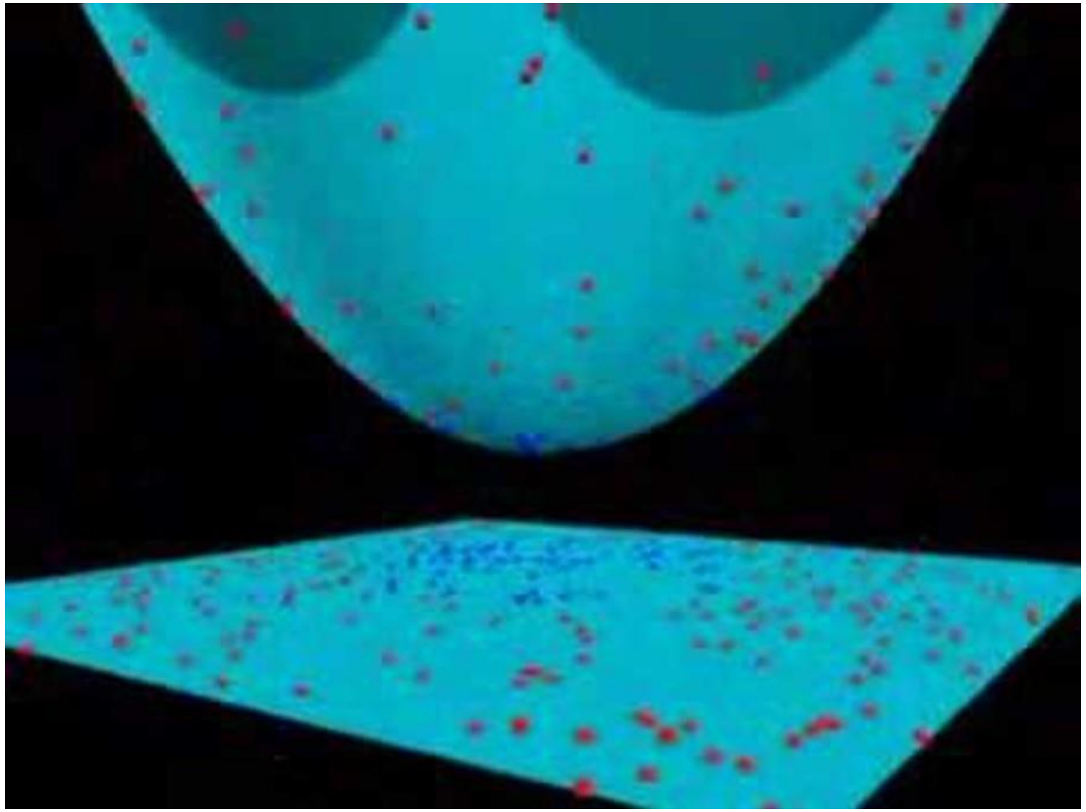
Maquinas de soporte vectorial NO lineales

Se puede mejorar a este clasificador lineal aplicando el “truco del kernel”

El resultado del algoritmo es similar, excepto que cada producto punto es reemplazado por una función kernel no lineal



Maquinas de soporte vectorial NO lineales



http://www.youtube.com/watch?v=3liCbRZPrZA

$$\max.\ W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

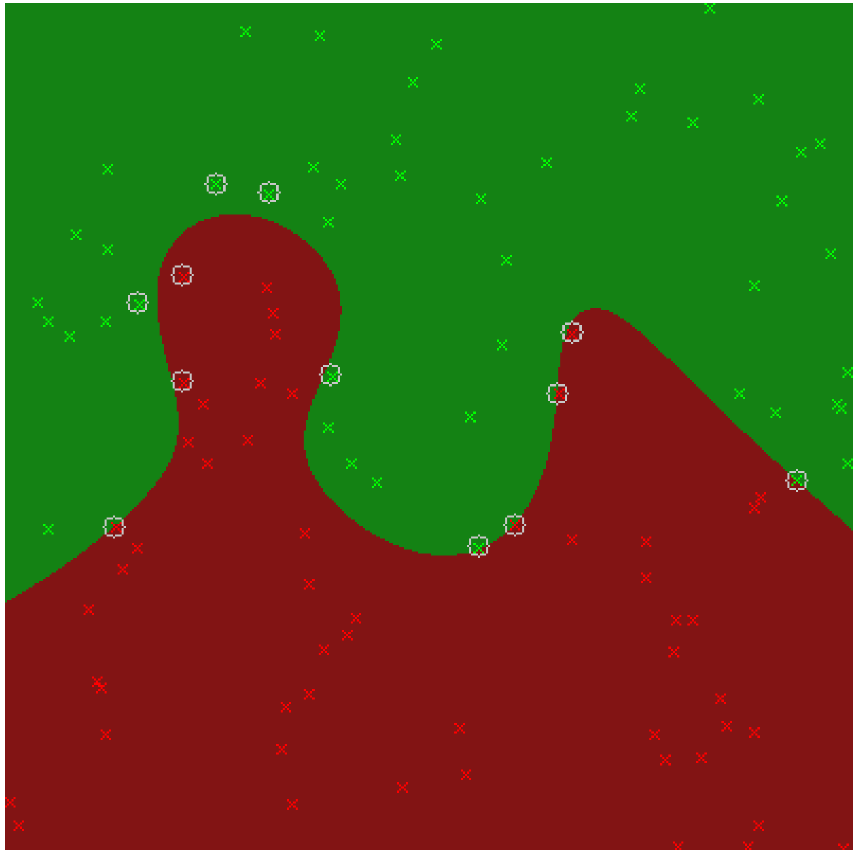
$$\Downarrow$$

$$\max.\ W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

Maquinas de soporte vectorial NO lineales

Polinomial (in homogéneo):

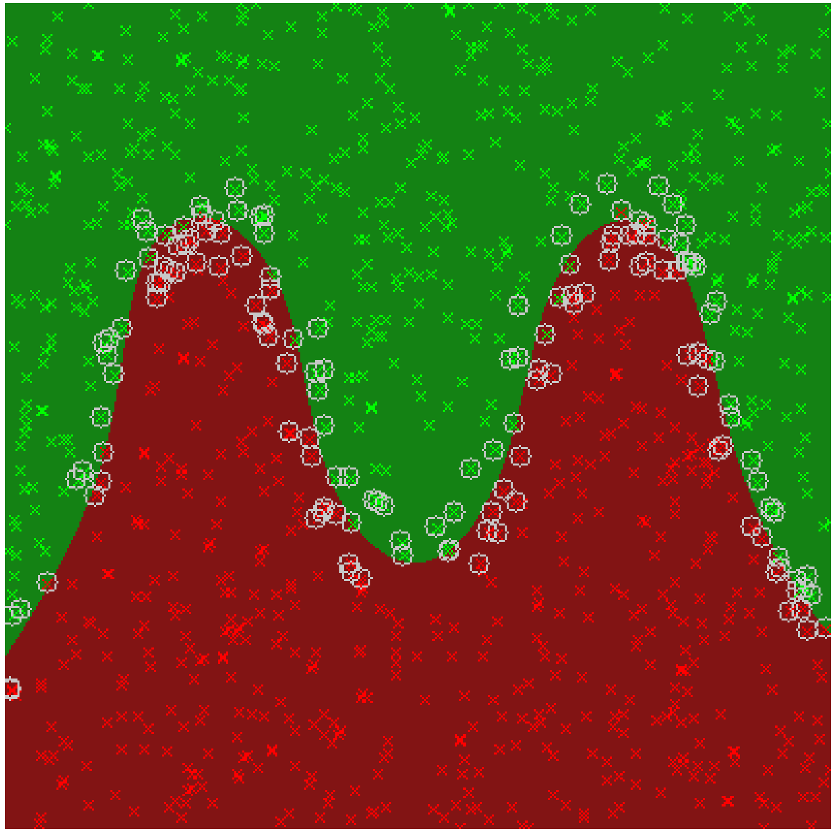
$$k(\mathbf{x},\mathbf{x}') = (\mathbf{x}\cdot\mathbf{x}'+1)^{d}$$



Maquinas de soporte vectorial NO lineales

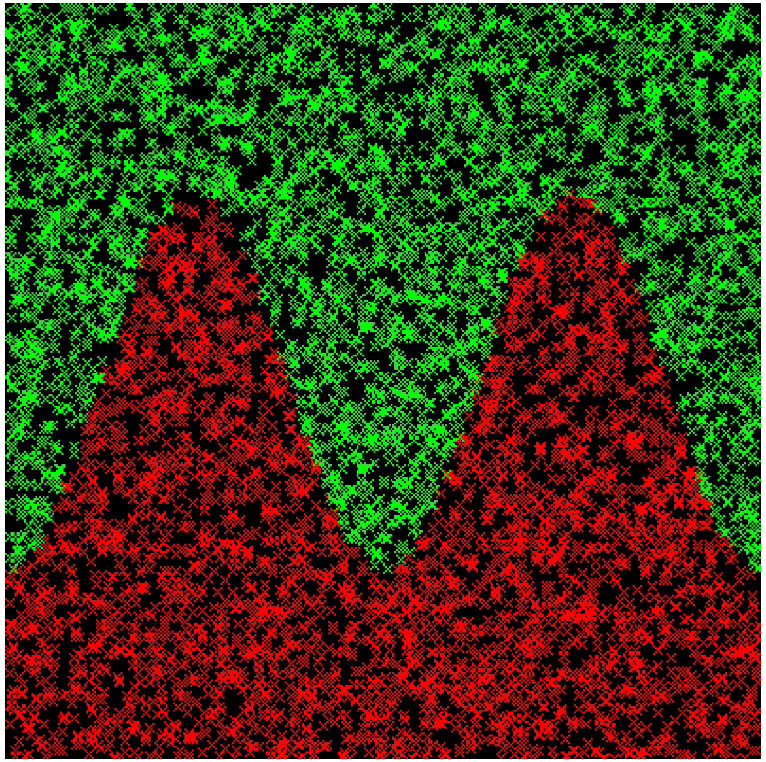
Función de Base Radial

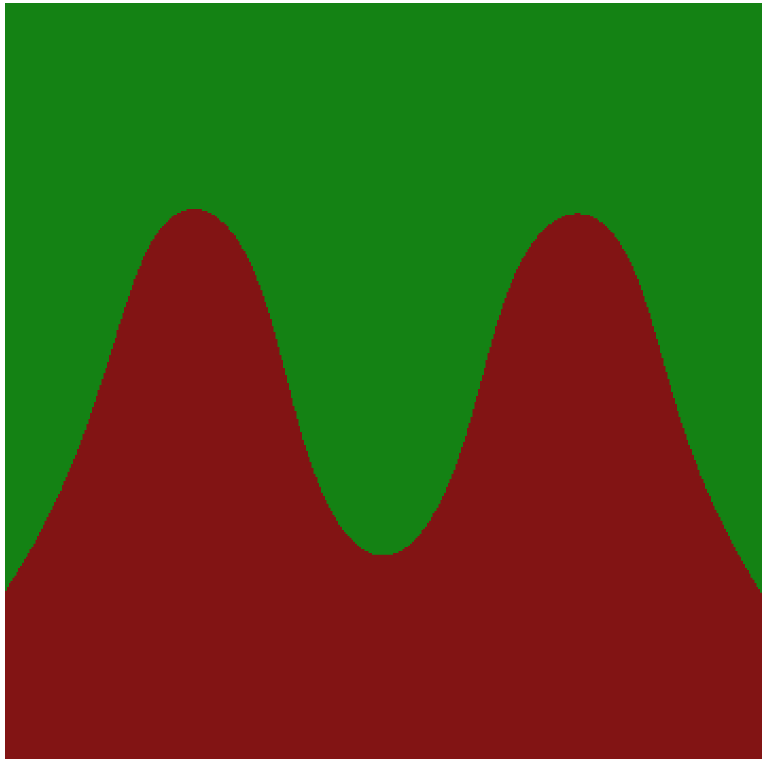
$$k(\mathbf{x},\mathbf{x}') = e^{(-\gamma \|\mathbf{x}\cdot\mathbf{x}'\|^2)} \text{ para } \gamma > 0$$



Maquinas de soporte vectorial NO lineales

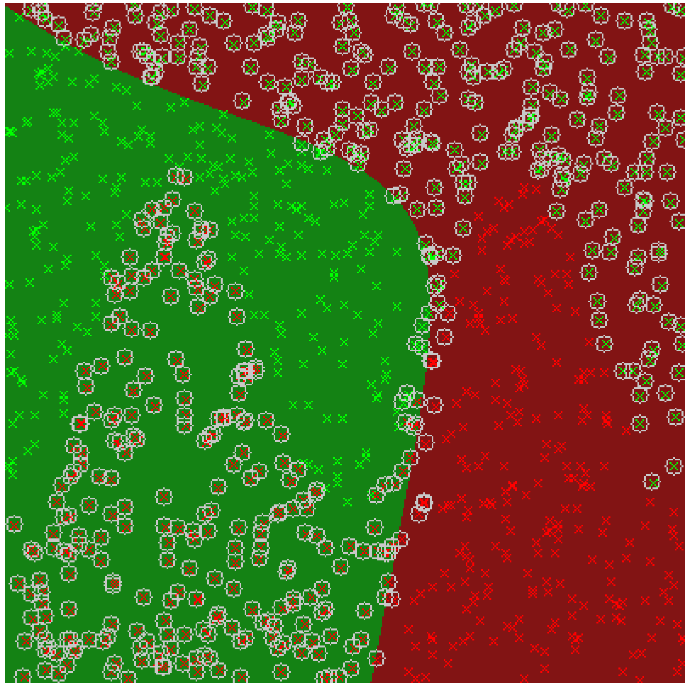
Función de Base Radial $k(\mathbf{x},\mathbf{x}') = e^{(-\gamma\|\mathbf{x}\cdot\mathbf{x}'\|^2)}$ para $\gamma > 0$

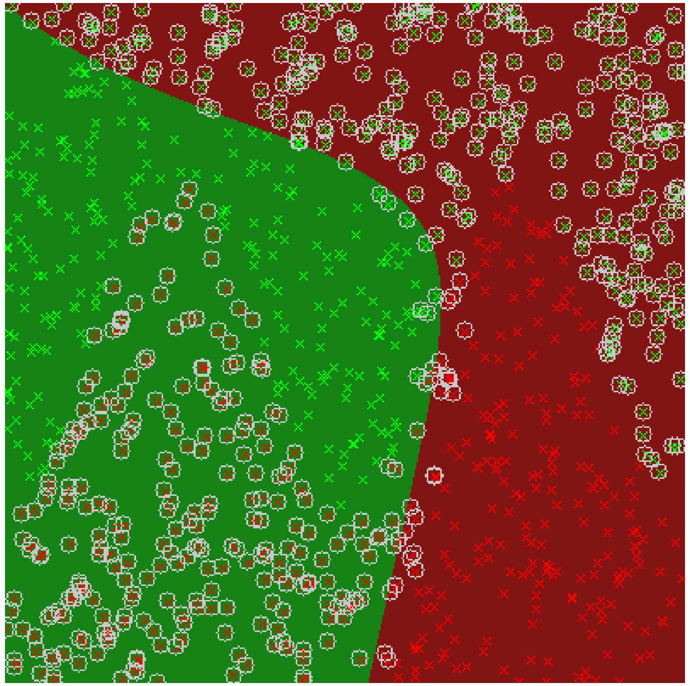




Maquinas de soporte vectorial NO lineales

Tangente Hiperbólica $$k(\mathbf{x},\mathbf{x}') = \tanh(\gamma \mathbf{x} \cdot \mathbf{x}' + c) \text{para algunos } \gamma > 0 \text{ y } c > 0$$





Bibliografía

- En memoria del profesor Patrick Winston. https://en.wikipedia.org/wiki/Patrick_Winston
- Corinna Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, 20, 1995. Disponible en http://homepages.rpi.edu/~bennek/class/mmld/papers/svn.pdf
- Christopher J. C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition". Data Mining and Knowledge Discovery 2:121–167, 1998. Disponible en https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/svmtutorial.pdf
- A Library for Support Vector Machines aplicación en línea http://www.csie.ntu.edu.tw/~cjlin/libsvm/
- 16. Learning: Support Vector Machines https://www.youtube.com/watch?v=_PwhiWxHK8o
- CVM Applet. aplicación en línea http://www.eee.metu.edu.tr/~alatan/Courses/Demo/AppletSVM.html
- Bases para el código de ejemplo en Opencv tomado de http://opencv.jp/sample/svm.html “en japonés”
- Support Vector Machine Disponible en http://en.wikipedia.org/wiki/Support_vector_machine
- SVM with polynomial kernel visualization disponible en http://www.youtube.com/watch?v=3liCbRZPrZA
- Excelente video resumiendo el concepto: https://www.youtube.com/watch?v=efR1C6CvhmE